

## Kuidas uurida esimese ja teise keele omandamist?

PILLE ESLON, KATRE ÕIM,  
ANNEKATRIN KAIVAPALU,  
REILI ARGUS, ERIKA MATSAK  
Tallinna Ülikool

**Ülevaade.** Artikli eesmärk on tutvustada esimese ja teise keele omandamist käsitleva uurimuse ideed. Esiplaanil on Eesti vajadused ning uurimistulemuste rakendatavus eesti keele õppes, tähelepanu keskmes esimese ja teise keele omandamise protsessi puudutavad küsimused: kuidas keeleoskus areneb, missugused keeleüksused (morfosüntaktilised konstruktsioonid, kollokatsioonid, idiomaatika jm) iseloomustavad keeleoskuse tasemeid. Kuna eesti keele põhjal ei ole esimese ja teise keele omandamist korpuslingvistiliselt ja võrdlevalt seni käsitletud, siis põhjendatakse vajadust taolise uurimuse järele, samuti selgitatakse, missugustest teoreetilistest ja metodoloogilistest alustest oleks seejuures mõttekas lähtuda, missuguseid analüüsimeetodeid kasutada. Artiklis kirjeldatava uurimissuuna arendamisega on alustatud Tallinna Ülikooli eesti keele ja kultuuri instituudis.

**Võtmesõnad:** korpuslingvistika; esimese ja teise keele omandamine; keeleoskustasemete lingvistiline kirjeldus; morfosüntaks; eesti keel

## 1. Kavandatava uurimuse olulisus

### 1.1. Eesti vajadused

Eestis nagu kogu Euroopa Liidus kasutatakse keeleoskustasemete kirjeldust, mille põhiseisukohad on sõnastatud “Euroopa keeleõppe raamdokumentis” (2007; edaspidi raamdokument). Viimase paari aasta jooksul on meil ilmunud raamdokumendi üldsätetele vastavad A, B ja C1 *can do*-stiilis tasemekirjeldused (Ilves 2008, 2010; Hausenberg jt 2008; Kerge 2008). Kõrvuti Prantsusmaal, Saksamaal, Soomes ja Lätis koostatud samalaadsete tasemekirjeldustega, samuti “Eesti keele arendamise strateegiaga 2004–2010” (2004) moodustavad need dokumendid alusmaterjali, millest on lähtunud põhikooli ja gümnaasiumi uues riiklikus õppekavas. Kuid ei raamdokumentis ega algaja, iseseisva või vilunud keelekasutaja kirjeldustes pole tasemeoskuste lingvistilist sisu tegelikult avatud, sest me ei tea täpselt, missugused keelestruktuurid ja sõnavara missugusele tasemele tegelikult omased on. Niisiis tuleb kõigepealt leida vastus küsimusele, kuidas õppija esimest ja teist keelt tegelikult omandab, mis seda protsessi iseloomustab, missugune lingvistiline sisu tagab ühelt keeleoskustasemelt teisele ülemineku jne. Ka õppeprotsessi korraldamise seisukohalt on oluline teada, mida inimene oma suhtluseesmärkide täitmiseks peab oskama. Eestis vastava uurimistöoga süstemaatiliselt ei tegelda, puuduvad korpuspõhised ja pikiuuringud, mille tulemusena saaks kirjeldada näiteks igale keeleoskustasemele iseloomulikke morfosüntaktilisi konstruktsioone. Ka sõnastike (sh õppe-sõnastike) koostamisel ei tugine ta korpusetele. Samas on näiteks inglise keele korpusi kasutatud nii sõnastike ja õpikute kui ka pedagoogiliste grammatikate koostamisel (vt Myles 2007: 377–400; Kilgariff 2003; Nielsen 2002; Summers 1996: 260–266). Küsimusele, kui palju sõnu teise keele õppija vajab, ei ole võimalik vastata, kui me ei tea, missugust osa kirjakeele sõnavarast valdab esimese keele ehk emakeelekõneleja, missugust osa sellest ta igapäevasuhtluses tegelikult kasutab, kui suure osa see tekstis keskmiselt katab (vt Nation & Waring 1997). Soome keeleteadlased on nendele küsimustele andnud vastuse juba 1979. aastal, kui ilmus

Oulu Ülikoolis koostatud soome keele sagedussõnastik: piisab 12 663 sõnast, et aru saada 90% tekstist (Saukkonen jt 1979: 7). Sõnastik leidis kohe laialdast kasutust soome keele õppes<sup>1</sup>. Sama ei saa väita 2002. aastal valminud eesti keele sagedussõnastiku (Kaalep & Muischnek 2002) kasutamise kohta, ehkki sõnastiku loomisele andis tõuke õpikutekstide keerukuse määramisel tehtud töö (vt Asser jt 2004).

Kuigi eesti keelt on riigikeelena õpetatud ja keeleõpet vastavalt korraldatud peaaegu kakskümmend aastat, ei rajane see töö teaduslikult põhjendatud alustel. Seda tõdeti järjekordselt põhikooli ja gümnaasiumi uue riikliku õppekava koostamisel, sest võõrkeelte ja eesti keele kui teise keele tasemeoskusi iseloomustava lingvistilise sisu kirjeldamisel tuleb jätkuvalt tugineda praktilisele õpetamiskogemusele ja üldteadmistele keele süsteemist ning struktuurist. Seetõttu on ka põhikooli ja gümnaasiumi uue riikliku õppekava võõrkeelteteadmisi kirjeldatud kooliastmeti, sõnaliikide ja kategooriate kaupa, sidumata neid kindlate keeleoskustasetega. Samas vajaksid nii algaja keeleõpetaja kui ka iseseisev õppija kindlat ettekujutust, mida tuleb osata, et sooritada tasemetest ja leida töö, mida saavutatud keeleoskustase teha lubab.

Mujal Euroopas on seda probleemi selgelt teadvustatud, ehkki tähelepanu keskmes on seni olnud keeleoskuse mõõtmine kui täiesti eraldi seisev uurimisala, millel pole kuigivõrd tegemist keele omandamise uurimisega<sup>2</sup>. Tasemeoskuste lingvistilise sisu empiiriliste uurin-gute koordineerimise eesmärgil on näiteks Charles J. Aldersoni, Jan

<sup>1</sup> Tänu prof Helena Sulkalale saame viidata ajalehes Helsingin Sanomat 6. aprillil 1979 ilmunud artiklile "Suomen kielen ensimmäinen frekvenssisanakirja valmis", milles kirjutatakse, et Londoni ülikooli professor Michael Branch andis välja soome keele sagedussõnastiku 2000 sagedasemat sõna tõlgituna kuude keelde (inglise, prantsuse, saksa, ungari, vene ja rootsi) ning lektor Hannele Branch kasutas seda sagedussõnastikku soome keele õpetamiseks Londoni ülikoolis. Aasta möödudes olid õppijad võimalised lugema Sillanpää teoseid.

<sup>2</sup> Vt European Association of Language Testing and Assessment (<http://www.ealta.eu.org/>), Language Testing Research Center (<http://www.ltrc.unimelb.edu.au/>), ajakiri Studies in Language Testing (<http://www.cambridgeesol.org/what-we-do/research/silt.html>) jt.

H. Hulstijni jt initsiatiivil moodustatud 12 Euroopa ülikooli baasil mitteformaalne SLATE-võrgustik<sup>3</sup>, et üheskoos leida vastus küsimusele, kuidas teise keele oskus areneb, milliseid konstruktsioone ja missugust sõnavara erinevate suhtluseesmärkide realiseerimiseks eri keeleoskustasemetel kasutatakse (vt Alderson 2006, 2007; Alderson & Huhta 2005; Hulstijn 2007). SLATE-võrgustikuga on ühinenud soome keele omandamise uurijad Jyväskylä Ülikoolist ning 2009. a kevadel osalesid töökoosolekul esimest korda ka Tallinna Ülikooli eesti keele ja kultuuri instituudi teadlased.

Teine oluline aspekt, mida keele omandamise uurimisel tuleks silmas pidada, on keeleõppija ealised iseärasused. Raamdokument kirjeldab täiskasvanute, mitte aga koolieelikute, noorema või keskmise kooliastme laste keeleoskust. Kuna osa eesti keele eksami sooritajatest on ikkagi põhikooli- ja gümnaasiumilõpetajad, siis on reaalne vajadus välja selgitada, kuivõrd edukalt saab täiskasvanutele mõeldud tasemekirjeldusi rakendada nooremate keeleõppijate puhul. Esmalt oleks vaja veenduda, kui sarnane või erinev on täiskasvanute ning koolieelikute, põhikooli ja gümnaasiumi õpilaste esimese ja teise keele omandamise protsess. Selle alusel saab võrrelda emakeelse õppija ja sama keelt teise keelena õppija keelelist arengut, välja tuua keelekasutuseelistusi ning keele omandamise universaalseid ja spetsiifilisi jooni (vt Ringbom 2007; Genoz jt 2001; Laufer 2000). Ka seda keeleomandamise aspekti pole eesti keele alusel süsteemselt uuritud ega kirjeldatud.

Siiani ei ole päris selge, kas suhtluses aktiivselt kasutatavad leksikaalsed üksused ja morfosüntaktilised konstruktsioonid muutuvad keeleõppija tasemeoskuste edenedes ühelt keeleoskustasemelt teisele järjest keerulisemaks ja mitmekesisemaks või mitte, kas nimetatud üksuste hulk suureneb, stabiliseerub või väheneb, kas väljendusvõimlused avarduvad või ei pruugi täiskasvanu keelekasutus oluliselt erineda teismelise omast (vt Bayley & Preston 1996). CEFLING-projekti “Linguistic Basis of the Common European Framework for L2 English and

<sup>3</sup> European SLATE network (Second Language Acquisition and Testing in Europe).

L2 Finnish (2007–2009)” inglise ja soome keele omandamise uurimise esialgsed tulemused (Martin jt 2010) näitavad, et kui B1-tasemel suureneb morfosüntaktiliste konstruktsioonide kasutussagedus, siis B2-tasemel suureneb samade keelendite kasutamise grammatiline korrektsus. Seda hüpoteesi on kavas testida Tallinna Ülikooli eesti keele ja kultuuri instituudis kavandatavas uurimuses. Teadlasi on huvitanud ka, milliseid etappe läbib õppija teksti produtseerides (vt Petch-Tyson 1998: 107–118), mis seda protsessi iseloomustab (nt Granger 1998; Ringbom 1998: 41–52; Cobb 2003: 393–423), kuidas ja missuguses järjekorras inimene sõnu ja vorme valib (nt Sinclair 2004) ning kasutama hakkab, missugune on keeleüksuste kasutusgrammatika (nt Tono & Aoki 1998).

Samuti pakub huvi keeleomandamise seotus loogiliste printsiipidega: kas teise keele omandamine tugineb samale arenguskeemile nagu väikelaste esimese keele omandamine; mis järjekorras võtab esimese ja teise keele õppija kasutusele loogilisi operaatoreid ja kvantoreid (vt Matsak 2009a, 2009b); mil määral sõltub väljendusoskus ehk teatud liiki loogiliste konstruktsioonide ning loogiliste tuletusreeglite rakendamine juba omandatud keeleoskusest ja individuaalsest võimekusest.

Kuna raamdokumendist on saanud Euroopa võorkeelepoliitika kõige mõjukam suunaja, siis on uutel teadmistel keeleoskustasemetega reaalse lingvistilise sisu kohta riiklikult oluline rakenduslik väljund. Eestis on see otseselt seotud Eesti keele arendamise strateegia ning muukeelse elanikkonna ja uusimmigrantide keeleõppega, eriti aga põhikooli ja gümnaasiumi riiklike õppekavade ning riigikeele tasemeõppe korraldamisega. Uue õppekava koostamine on teravalt esile toonud vajaduse saada täpselt teada, missugune peaks olema esimese ja teise keele tasemeoskuste tegelik areng ning lingvistiline sisu. See aitaks mõista keeleomandamisprotsessi olemust. Põhjalikud teadmised eri vanuses ja erineva keeleoskusega esimese ning teise keele õppijate tegelikust keelekasutusest loovad eelduse emakeele ja teise keele õppe meetoodika täiustamiseks ning nende senisest tunduvalt tihedamaks integreerimiseks. See peaks olema aineprogrammide, õpikute ja õppematerjalide ning keeleoskuse mõõtmise meetoodika ja õpetajakooolituse täiustamise alus.

## **1.2. Pilguheit eesti keele omandamisele esimese ja teise keelena: lingvistiline aspekt**

Keeleõppe seisukohalt pakuvad meile huvi esimese keele käsitlused sellest, kuidas väikelaps emakeelt omandab (nt Argus 2008a; Vija 2007; Vihman & Vija 2006) ja kuidas laps hakkab keeleoskuse arenedes loogilisi konstruktsioone kasutama (Matsak 2009a, 2009b). Vähem süsteemselt on tegeldud koolilaste ning täiskasvanute keeleomandamise ja keeeloomega. Küll aga on seda protsessi uuritud lingvistilisest aspektist (nt Vare 2001; Vill 2001; Torn 2003, 2004; Paomees 2003; Jüssi 2005; Saareväli 2005; Lepajõe 2006; Habicht 2006; Jürine 2007 jt).

Analoogne olukord valitseb ka eesti keele kui teise keele uurimises. Huvi on pakkunud mitmed konkreetset lingvistilised küsimused (vt Hausenberg 2003; Metslang jt 2003; Alp 2004; Vanem 2006; Kitsnik 2006, 2007; Pool 2007; Pastuhhova 2007 jt), kuid puudub süsteemne õppijakeele analüüs, nagu ka esimese ja teise keele omandamise protsessi korpuspõhine võrdlus.

Samas on mitmete teiste õppijakeelte, nt inglise, saksa, prantsuse, hispaania, taani, rootsi, norra jne põhjal uuritud õppijakeele kollokatsioone, sõnade distributiivseid omadusi, semantilisi ja morfoloogilisi eelistusi, tüüpilist/ebatüüpilist sõna- ja vormikasutust, süntaktilisi konstruktsioone, diskursust, stiilivariante jne (vt Sinclair 2004; Granger 1997; 1998; Biber jt 1998; Stubbs 2001; Tognini-Bonelli 2001; Leech jt 2001; Reppen jt 2002; Biber 2004; Nesselhauf 2005; Römer 2005, 2007; Jantunen 2004, 2008 jt). Nendes õppijakeele uurimustes on kasutatud erinevaid korpusainestikke, korpuslingvistilisi lähenemissuundi, kvantitatiivseid analüüsimeetodeid, keeletarkvara ja standardprogramme.

## **1.3. Kavandatava uurimuse eesmärgid ja rakenduslik väljund**

Uurimuse põhieesmärk on võrrelda esimese ja teise keele arengut kahes suunas: ühelt poolt vanuserühmade kaupa ehk etapiti (väikelaps – koolieelik – põhikooliõpilane – gümnaasiumiõpilane – täiskasvanu) ja teisalt

keeleskustasemeti (A1 – A2 – B1 – B2 – C1 – C2). Nende koordinaatide ristumispunktid moodustavad raamistiku, mille alusel saab hakata kirjeldama esimese ning teise keele kirjalikku ja suulist produktsiooni.

Kõigepealt tehakse kindlaks iga etappi ja saavutatud taset iseloomustav sõnavara (k.a. idiomatilised üksused) ning vaadeldakse sõnavormide kasutamist teatud liiki morfosüntaktilistes konstruktsioonides, mis omakorda on seotud kindlat tüüpi kontekstide produtseerimisega. Iga keeleskustaset iseloomustavat sõnavara ja grammatikastruktuure on kavas kirjeldada nende sageduse, grammatilisuse ning aktsepteeritavuse alusel. Seeläbi saadakse vastus küsimusele, mil määral tugineb teise keele õppimine reeglitele, mil määral analoogiale. Samuti fikseeritakse keeltevahelisest mõjust ja ülekandeprotsessidest tingitud arengud.

Kuna peetakse tõenäoseks, et keeleüksused ilmuvad õppija keelde kindlas järjekorras ja et esimese keele omandamisel toetutakse peamiselt keeleomastele mitmesõnalistele idiomatilist laadi üksustele ning teist keelt õppides emakeele ja õpitava keele vahelistele analoogiaseostele, siis on kavas seda seisukohta kontrollida. Sel eesmärgil võrreldakse, kuidas esimese ja teise keele oskus erinevatel etappidel areneb, sidudes selle raamdokumendis sätestatud keeleskustasemete lingvistilise sisu kirjeldamisega.

Kuna esimese ja teise keele omandamise etappide ning raamdokumendis sätestatud keeleskustasemete vahel pole üksühest vastavust, siis eeldame järgmist. Mida on raske omandada mistahes keeles, on suure tõenäosusega raske omandada ka eesti keeles<sup>4</sup> (nt sõnajärg, rektsioon, ühildumine). Mis on raske esimese keele omandaja jaoks ning saadakse selgeks ajaliselt hiljem, on suure tõenäosusega raske ka teise keele õppija jaoks. Esimese ja teise keele omandamise loogilised skeemid on ilmselt sarnased ning teist keelt õppides areneb keeleskus samamoodi kui esimest keelt omandades, kusjuures mõlemal juhul konstrueeritakse lauseid vanusele vastavate loogiliste implikatuuride kohaselt.

---

<sup>4</sup> Vt <http://www.zas.gwz-berlin.de/cost/downloads.html> (12.09.2008) – *memorandum of understanding*.

Esimese ja teise keele võrdleva lingvistilise uurimise tulemusena saadakse kõigepealt ülevaade standard- ehk kirjakeele sagedastest vormidest ja konstruktsioonidest, mida emakeelekõneleja on eelistanud kasutada oma suulises ja kirjalikus produktsioonis. Neid tulemusi võrreldakse eesti õppijakeeles sagedaste vormide ja konstruktsioonidega. Eesmärk on keskenduda verbi argumendistruktuuridele, reksioonimallidele, grammatilistele vormidele jm, mida emakeelekõneleja ja teise keele õppija produtseerivad samalaadselt, erinevalt või analoogia põhjal. Selleks on kavas vaadelda ajalis-aspektilise ja modaalse diskursuse tüüpilisi/ebatüüpilisi semantilisi ja grammatilisi piiranguid ning sünonüümset varieerumist.

Uurimistulemustel on oluline rakenduslik väljund: kujundada teaduslikult põhjendatud alus tasemeõppe korraldamiseks ja keeleoskuse mõõtmiseks, tasemeoskuste saavutamisele suunatud õppesõnastike ja õpikute koostamiseks. Kuna uuringu käigus analüüsitakse kõikide vanuserühmade keelekasutust väikelastest täiskasvanuteni, siis on uurimistulemused hõlpsasti võrreldavad samalaadsete rahvusvaheliste COST-projekti ja SLATE-võrgustiku raames tehtud uuringute tulemustega. Uurimistöö teoreetiline tähtsus seisneb panuses esimese ja teise keele omandamise teooriasse, mis on eesti keele seisukohalt uuenduslik.

## 2. Uurimisobjekt ja analüüsisuunad

### 2.1. Uurimisobjekt ja allikad

Võrdleva korpusuuringu objektiks on kaks eesti keele kasutusvarianti – eesti keel kui esimene keel (emakeel) ja eesti keel kui teine keel (eesti õppijakeel). Keeleanese allikad on elektroonsed korpused: eesti lapsekeele andmebaas<sup>5</sup> ja Tallinna Ülikooli eesti vahekeele korpus (EVKK)<sup>6</sup>.

<sup>5</sup> Vt CHILDES, <http://childes.psy.cmu.edu/data/Other/> (29.08.2009).

<sup>6</sup> <http://evkk.tlu.ee>; vt ka <http://www.keeletehnoloogia.ee/projektid/vako/vako> (18.09.2009).

Võrdluskorpustena saab kasutada Tartu Ülikooli eesti kirjakeele korpust (TÜKK)<sup>7</sup> ja Eesti Keele Instituudi tekstikorpust (EKI korpus)<sup>8</sup>. Uudsenda tuleb käiku Eesti fraseologismide elektrooniline alussõnastik<sup>9</sup> (vt K. Öim 2009), mida saab rakendada esimese ja teise keele idiomaatilise võrdlemiseks. Kuna see sõnastik on koostatud mõisteseoste alusel ja morfosüntaktiliselt märgendatud, siis aitab see fraseologisme tekstis paremini tuvastada. Õppijakeele keeletehnoloogilist ressursi on vaja suurendada ning selle tööga tegeldakse: uuendatakse Eesti vahekeele korpuse struktuuri, lisatakse allkorpusi, mis sisaldavad Riikliku Eksami- ja Kvalifikatsioonikeskuse (REKK) üleriigilisi tasemetöid, põhikooli ühtlustatud küsimustega eksamitöid, riigieksamitöid, eesti keele tasemeeksami töid ning Eesti Vabariigi põhiseaduse ja kodakondsuse seaduse tundmise eksamitöid.

Niisiis on esimese ja teise keele võrdlevaks uurimiseks Eestis olemas erineva mahuga ning erinevatel põhimõtetel koostatud korpusi, mis sisaldavad mitmesugust päritolu keeleainest. Samas on rõhutatud, et võrreldavad valimid peavad vastama üksteisele nii mahult, materjali valikuprintsiipi kui ka korpuse metainfo poolest (vt Biber 1993, 2004; Baker 2006), sest vaid sel juhul saab uurimistulemusi pidada usaldusväärseteks. Niisugust ideaalselt tasakaalustatud eesti keele ressursi, kus oleks samaväärselt esindatud eesti keele kui esimese ja teise keele kasutusnäited ning suuline ja kirjalik tootmine, praegu ei ole. Seetõttu võetakse keekekasutuseeliste ning esimese ja teise keele sarnasuste/erinevuste väljatoomise aluseks võrreldavate valimite statistiline olulisustõenäosus (vt Kilgarriif 2001; Eslon & Matsak 2009: 81–84, 86–89).

<sup>7</sup> <http://www.cl.ut.ee> (30.10.2009).

<sup>8</sup> <http://www.eki.ee/corpus/> (3.09.2009).

<sup>9</sup> <http://www.keeletehnoloogia.ee/projektid/fraseologismid/eesti-fraseologismide-elektroonilise-alussonastiku> (13.09.2010).

## 2.2. Analüüsisuunad

Korpusi on keelekasutuse uurimiseks rakendatud erinevalt, tavaliselt allikmaterjalina mingite teoreetiliste seisukohtade illustreerimiseks, mis näitab uurimuse orienteeritust teatud lingvistilise koolkonna teorialele (vt Leech 2004: 61). Seda eesmärki teenib korpuspõhine keeleanalüüs (ingl *corpus-based language analysis*; sünonüümselt ka hüpoteesist tulenev keeleanalüüs ehk *hypothesis-driven analysis*, vt Rayson 2002: 1). Uurijale huvi pakkuvate keelenähtuste leidmiseks korpusest rakendatakse erinevat laadi kvantitatiivset analüüsi, mille tulemusi tõlgendatakse lingvistiliselt (vt Rayson 2002: 13; Granger 2003).

Teine võimalus on korpusainesest tulenev uurimissuund (*corpus-driven research* või *data-driven research*), kus korpus pole pelgalt teoreetiliste seisukohtade tõestamiseks vajalik allikmaterjali kogu. Seda keeleuurimise korpuslingvistilist paradigmat on nimetatud ka andmetest tulenevaks uurimissuunaks (*data-driven paradigm characteristic of corpus linguistics*, vt Leech 2004: 61) – oluliseks peetakse teavet selle kohta, kuidas keelekasutuses on põimunud sünkroonne varieerumine ja diakroonilised muutused (vt Leech 2004: 77–78). Korpusest tuleneva lähenemise eelis korpuspõhise keeleanalüüsi ees on selles, et uurija ei otsi korpusdest mitte tõestusi oma hüpoteesi(de)le, vaid püüab interpreteerida seda, mis iseloomustab reaalselt keelekasutust ning võib seetõttu uurijale üldse huvi pakkuda (vt Tognini-Bonelli 2001, 2002; Hanston 2002; Granger 2004; vt ka diskursusanalüüsi alased uurimused nagu Stubbs 1996, Orpin 2005; Hoey jt 2007). Korpusainesest tulenevat analüüsisuunda rakendades saab leida keelekasutuses eelistatud ja erinevatele keelekasutusvariantidele omased leksikaalsed üksused ja morfosüntaktilised konstruktsioonid. Tuvastatakse sõna- ja vormisagedus ning aktsepteeritava/mitteaktsepteeritava keelekasutuse vaherkord. Analüüsiüksusteks on sõnavormid, leksikaalsed terviküksused ja morfosüntaktilised konstruktsioonid, mida sarnastel suhtluseesmärkidel on aktiivselt kasutatud. Kui aga uurida sagedaste keeleüksuste sünonüümset varieerumist, siis oleks sobivam jätkata korpuspõhise analüüsiga. Seetõttu on nii keele-

kui ka diskursusanalüüsis peetud oluliseks rakendada mõlemat uurimissuunda integreeritult (vt Behrens 2008: XXVII; Orpin 2005: 38–39), kombineerides samas erinevaid statistilisi, lingvistilisi, psühholingvistilisi, sotsiolingvistilisi analüüsimeetodeid (vt Taylor 2008: 183).

Et välja selgitada, missuguseid lingvistilisi mustreid erinevates tekstiliikides esineb ja missugusel suhtluseesmärkidel neid on eelistatud kasutada, peetakse korpuslingvistikas vajalikuks võrrelda erinevaid terviktekste ning testida saadud tulemusi suuremahuliste võrdluskorpuste peal, tavaliselt kirjakeele- ehk standardkorpustel (nt Stubbs 1996: 126–154). Seda suunda on nimetatud korpuste abil teostatavaks analüüsiks (*corpus-assisted analysis*) ja kasutatud näiteks õpikutekstide ja muude keeleõppematerjalide sõna- ning vormikasutuse autentsuse hindamisel (vt nt Römer 2007), registrierinevuste esiletoomiseks (nt Biber 2004) jne. Lisaks ükskeelsele korpusainesele on peetud vajalikuks kaasaata ka muid allikaid (paralleel- ja tõlkekorpused, veebimaterjalid, spetsiaalselt uurimiseesmärgil produtseeritud tekstid jne).

Kavandatavas esimese ja teise keele tasemeoskuste lingvistilise sisu määramise uurimuses rakendatakse korpusest tulenevat analüüsi, kombineerides seda vastavalt vajadustele korpuspõhise ja korpuste abil teostatava analüüsiga. Seega on tegu integreeritud korpuslingvistilise uurimissuunaga. Korpusest tuleneva pikiuuringuga oleks võimalik jõuda keeleomandamisetappide üldkirjelduseni, tuua välja harvad või raskemini omandatavad nähtused, et neid katsete ja tajutestidega uurida<sup>10</sup>. See annab keeleomandamisprotsessi kohta lisateavet ja võimaldab kontrollida mõningaid keeleõppes kehtivaid pedagoogilisi tavaaru-saamu (vt Argus 2009, 2008b). Esimese ja teise keele tekstiloome eripära on kavas fikseerida *ScriptLog*-programmi abil (vt Kaivapalu 2009) jm.

Keeleomandamisprotsessi loogika esiletoomiseks kasutatakse tehis-intellekti raames spetsiaalset välja töötatud dialoogsüsteemi DST (vt

<sup>10</sup> COST-projekti “Robust Stages in First Language Acquisition” raames, mille töös osaleb Reili Argus, on välja töötatud rida keeleomandamiskatseid, mille põhjal uuritakse aspektiliste situatsioonide tajumist ja sõnavormide kasutamist, fikseeritakse teksti produtseerimisega seotud nähtusi.

Matsak 2009a, 2009b), mis transformeerib loomuliku keele lause predikaatarvutuse valemiks. DST rakendamine esimese ja teise keele produtseerimise protsessi analüüsil võimaldab jälgida loogiliste konstruktsioonide järk-järgulist kasutuselevõttu, samalaadsust ja erijooni. Kuna DSTs kasutatakse eesti keele morfoanalüsaatorit, siis lubab see siduda keelendite morfoloogilise ja loogilise rolli lauses, leida loogiliste struktuuride esilduvust iseloomustavaid lingvistilisi mustreid ning ainu-kordseid kasutusi.

Esimese ja teise keele jaoks tüüpiliste/ebatüüpiliste sõnavormide, leksikaalsete terviküksuste, morfosüntaktiliste konstruktsioonide ja nende sünonüümse varieerumise uurimisel on oluline rakendada erinevaid statistilise analüüsi võimalusi, kuna see omab humanitaarvaldkonnas tervikuna heuristilist tähendust, aidates esile tuua keelenähtusi, mida tavapärase lingvistiline analüüs tuvastada ei võimalda (vt Atkins & Levin 1995; Abney 1996; Oakes 1998; Tooding 1999; Granger 2005 jt). Seetõttu on korpusanalüüsis laialdast rakendamist leidnud ka erinevad statistikapõhised programmid (vt Rayson 2002; Scott & Tribble 2006; Chambers 2007). Eesti keele kui esimese ja teise keele võrdlevas uuringus on kavas kasutada programmi WordSmith Tools, mille abil tuuakse esile sõnavormide sagedus, leitakse kollokatsioonilised üksused, saadakse ülevaade nende esilduvusest ja kasutuskontekstide varieerumisest. Järgnev automaatne süntaktiline analüüs võimaldab välja tuua erinevatele keelekasutusvariantidele omased morfosüntaktilised konstruktsioonid, mida võrreldavates valimites on kasutatud kaks ja enam korda.

Korpuslingvistilise suuna, keeletarkvara ning erinevate statistikute integreeritud ja järjekindel rakendamine keelekasutuse uurimisel omab lingvistika jaoks metodoloogilist tähendust (vt Lindquist & Mair 2004), kuna kasutuspõhisuse printsiip võimaldab siduda erinevaid teooriaid ja uurimisvaldkondi.

### 3. Kavandatava uurimuse teoreetiline tagapõhi

Eesti keele kui esimese ja teise keele omandamise protsessi võrdlemisel tuginetakse korpusandmetele ning selles mõttes on kavandatav uurimus olemuselt ja sisult kasutuspõhine. Erinevatesse korpusse koondatud keeleainese põhjal selgitatakse välja, kuidas eri vanuses ja erineva keeleoskustasemega inimesed eesti keelt kui emakeelt ja teist keelt kasutavad, kuidas nende keeleoskus samm-sammult edeneb (vt nt Tomasello 2006: 259; Wilson 2003: 83; Leistyna & Meyer 2003). Arvestada tuleks ka keelekasutuse individuaalse varieerumisega ja varem omandatud keel(t)e mõjuga iga järgneva keele omandamisele (vt Jarvis & Odlin 2000; Pavlenko & Jarvis 2002; Kaivapalu 2005; Kaivapalu & Martin 2007 jt), kuid keeleoskustasemete lingvistilise sisu kindlaks tegemisel on siiski kõige olulisem leida tasemele iseloomulikud keelekasutusmustrid.

Teooria tasemel võimaldab kasutuspõhisuse printsiip integreerida kolm olulist keeleomandamisteoreetilist lähtekohta, milleks on keele töödeldavusteooria (*Processability Theory*), ühtne võistlusmudel (*Unified Competition Model*) ja kontseptuaalne semantika. Allpool tutvustamegi lühidalt nimetatud teooriate põhiseisukohti.

#### 3.1. Keele töödeldavusteooria, ühtne võistlusmudel ja keeltevaheline mõju

Manfred Pienemanni töödeldavusteooria (*Processability Theory*, Piene-mann 1998, 2005) kohaselt ilmuvad sõnad, väljendid, vormid ja keelestruktuurid õppijakeelde kindlas järjekorras ning seda järjekorda on võimalik teatud määral ennustada. Tabelis 1 on kirjeldatud inglise keele omandamise astmelisust, mida iseloomustavad hierarhilisus ja lingvistiline varieeruvus. Keeleoskuse areng on jaotatud kuueks tasemeks, iga taseme lingvistiline sisu on piiratud kindlate väljendite, vormide ja keelestruktuuridega, millest õppija hakkab keeleoskuse arenedes moodustama keerukamaid morfosüntaktilisi konstruktsioone.

**TABEL 1.** *Inglise keele morfosüntaksi omandamise kuus taset*  
(Kefßler 2009)

Tase	Keelestruktuurid	Näited
6	Cancel Aux-2nd	I wonder <b>what he wants</b> .
5	Neg/Aux-2nd-? Aux-2nd-? 3sg-s-	Why <b>di-d ' - n 't</b> you tell me? Why <b>can 't</b> she come? Why <b>did</b> she eat that? What <b>will</b> you do? Peter likes bananas.
4	Copula S (x) Wh-copula S (x) V-Particle	<b>Is</b> she at home? <b>Where</b> is she? Turn it <b>off!</b>
3	Do-SV(O)-? Aux SV(O)-? Wh-SV(O)-? Adverb-First Poss (Pronoun) Object (Pronoun)	<b>Do</b> he live here? <b>Can</b> I go home? <b>Where</b> she went? <b>What</b> you want? <b>Today</b> he stay here. I show you <b>my</b> garden. This is <b>your</b> pencil. Mary called <b>him</b> .
2	S neg V(O) SVO SVO-Question -ed -ing Plural -s (Noun) Poss -s (Noun)	<b>Me no</b> live here. / I don 't live here. Me live here. You live here? John <b>played</b> . Jane <b>going</b> . I like cats. Pat 's cat is fat.
1	Words Formulae	Hello, Five Dock, Central How are you? Where is X? What 's your name?

Teooriana on Pienemanni teooria universaalne: seda mudelit saab rakendada mis tahes õppijakeele uurimiseks, sh esimese ja teise keele tasemeoskuste kirjeldamiseks ning iga uut taset iseloomustavate morfosüntaktiliste konstruktsioonide esiletoomiseks<sup>11</sup>. Samas on tõdetud, et

<sup>11</sup> Nimetatud seisukohtadest lähtuti ka Jyväskylä ülikoolis teise keele omandamisprotsessi ja keeleoskuse tasemelise arengu uurimiseks algatatud CEFLING-projektis "Linguistic Basis of the Common European Framework for L2 English and L2 Finnish (2007–2009)", vt <http://www.jyu.fi/hum/laitokset/kielet/cefling/en/subprojects>

germaani keelte põhjal välja töötatud teooria pole piisavalt detailne selgitamaks nt soome õppijakeele morfo(fono)loogilist arengut (vt Martin 2004). Oletatavasti kehtib sama eesti keele kohta.

Pienemanni teooria on universaalne ka teisest seisukohast: selle aluseks on kontseptuaalsemantiline arusaam teadmuse talletamisest pikaajalises mälus teadmismustritena. Keeletöötlusprotsessis see avaneb. Teadmismustrite kontseptuaalne semantiline hierarhia eksplitseerub olenevalt suhtlusvajadustest, ilmneb keele produtseerimisel ja keelelise väljenduslaadi valiku kaudu, kui semantiline kontsept seostatakse vastava predikaadi argumendistruktuuridega (vt ka H. Õim jt 2009: 492). Eriti kujukalt toimib see mudel teise keele õppija suulises produktioonis, mille puhul lingvistikateadmistel on teisejärguline tähendus. Ouline on leida kiiresti oma mentaalsest leksikonist mõtte edastamiseks vajalik semantiline kontsept, mis määrab mitte ainult lause argumendistruktuuri, intonatsiooni ja hääldusviisi, vaid sisaldab ka lingvistilist informatsiooni, millest oleneb nt aja, aspekti, isiku, arvu, objekti- ja subjektiäände valik. Seega on nii suulise ja kirjaliku teksti produtseerimisel kui ka mõistmisprotsessis kesksed niisugused lingvistilised kategooriad nagu verb, substantiiv, adjektiiv.

Keeleüksusi kasutatakse valivalt, teatud liiki morfosüntaktilistes kompleksides (keeleelementide modulaarne organiseeritus) ilmneb ühe või teise grammatilise vormi eelistamine. Esimese keele puhul toimub see nii, et teatud vormide kasutamine tingib teiste teatud vormide ilmlemise jne. Teise keele puhul tuleb niisuguste toimingute jada meelde jätta. See teadmus talletatakse lühimälus assotsiatiivse siduse alusel oma mentaalse leksikoniga, millest lähtudes hakatakse omandatud mõtete väljendamiseks rakendada. Lühimälu aluseks olevad kontseptid on üldised ja hägusad, assotsiatsiooniseosed mentaalse leksikoniga

---

(16.11.2009). Tööd jätkatakse Topling-projektiga “Paths in Second Language Acquisition (2010–2013)”, mille eesmärk on välja selgitada, kuidas soome keele kui teise keele ning inglise ja rootsi keele õppijate kirjalik väljendusoskus edeneb. Aluseks on koolilaste pikiuuringu tulemused, mida võrreldakse CEFLING-projektis saadud tulemustega, vt <http://www.jyu.fi/hum/laitokset/kielet/topling> (20.03.2010).

võivad olla (ligi)lähedased või ekslikud. Samas rajaneb teise keele õppimine ja omandamine just nendel hägusatel, ent universaalsetel kontseptidel ja assotsiatsioonidel.

Arusaam keelemärgist kui vormi ja tähenduse ühtsusest ning keeletöötlusprotsessist kui tähenduse keelendamise vormi abil on ka Brian MacWhinney (2004) ühtse võistlusmudeli aluseks. MacWhinney seisukohast on nii esimese kui teise keele omandamine mõistetav selle keele süsteemi üles ehitamisena<sup>12</sup>. Kuigi laste esimese keele omandamine erineb täiskasvanute teise keele omandamisest vähemalt kolmes aspektis (maailma tundma õppimine samaaegselt keele omandamisega, paindlikum ajutegevus, hooldajate sotsiaalne tugi), on nii esimese kui teise keele omandamise meetodid ja strateegiad samad ning mõlemad protsessid omavahel tihedalt põimunud. MacWhinney arvates kinnitavad seda vähemalt kaks teise keele õppimisega kaasnevat nähtust: keeleülekanne (eriti analoogiale tuginev positiivne transfer) ja koodivahetuse mehhanismide rakendumine. Lapsed on teatud kriitilise vanusepiirini lisaks esimesele keelele avatud ka teis(t)e keel(t)e suhtes: sel perioodil toetub emakeele omandamine ja mitmekeelsuse kujunemine samadele alustele. Ajaliselt hiljem, täiskasvanuna hakatakse teise keele õppimise käigus õpitava keele süsteemi enda jaoks uuesti üles ehitama, st assotsiatsioonide alusel oma mentaalse leksikoniga luuakse uus, õpitava keele produtseerimiseks mõeldud protsessisüsteem, millele tuginedes aktiveeritakse sõnavara, valitakse vastav morfosüntaktiline vormistus ning kavandatakse intonatsioonistruktuur ja hääldusviis.

Teise keele omandamise protsessi mõjutavad paljud, kohati vastandlikud ja erisuunalised tegurid, millest olulisim on esimese keele ja teiste varem omandatud keelte mõju. Kuigi keeltevahelise mõju uurimisest on välja kasvanud iseseisev uurimisvaldkond, on selle suuna põhiseisukohad olulised, et mõista esimese ja teise keele tasemeoskuste kujunemise psühholingvistilist mehhanismi.

<sup>12</sup> Vt Brian MacWhinney, <http://psyling.psy.cmu.edu/papers/CM-general/unified.pdf> (4.03.2010).

Keeltevahelise mõju uurimisel tuleb vahet teha objektiivsel, tunnetatud ja oletataval sarnasusel (vt Ringbom 2007: 7–9, 24–26). Tegelik ehk objektiivne sarnasus (*actual/objective similarity*) kuulub lingvistika pädevusse ning seda on võimalik teoreetiliselt määratleda keelesüsteemide võrdluse põhjal (nt eesti ja soome morfoloogiasüsteemide võrdlus). Keeltevaheline tegelik sarnasus on sümmeetriline. Tunnetatud sarnasus (*perceived similarity*) seevastu ei pruugi olla tingimata sümmeetriline. See, kuidas õppija keelesüsteemide sarnasust tajub, on tegelikust keeltevahelisest sarnasusest tunduvalt avaram psühholingvistiline nähtus ja seetõttu on tunnetatud sarnasusest rääkides sageli kasutatud termineid *psühhotüpoloogia* ja sarnasuse *psühholingvistiline reaalsus* (vt Kaivapalu 2005). Kuna tunnetatud sarnasus sõltub paljuski õppijate individuaalsetest iseärasustest, siis on seda raskem määratleda. Keeleloomes, eriti emakeelest tüpoloogiliselt erineva keele puhul, on olulisem oletatav sarnasus (*assumed similarity*), mis kuulub nagu tunnetatud sarnasuski psühholingvistika valdkonda. Oletatav sarnasus võib põhineda ja sageli põhinebki tunnetatud sarnasusel, kuigi ka see ei tarvitse alati nii olla. Kui õppija ei tunneta sarnasusi emakeele ja õpitava keele vahel, oletab ta, et viimane toimib ligikaudu samamoodi kui emakeel või teised varem omandatud keeled. Kui õppija oletused ei pea paika, siis põhjustab see keeleloomes vigu: emakeele mõju ilmneb otseselt sõnade, väljendite, vormide ja keelestruktuuride valikus ning nende kombineerimises. Keeleülekanne on jälgitav nii keeleloome- kui ka mõistmisprotsessis (Möhle & Raupach 1989: 202; Kaivapalu & Martin 2007: 132–133).

Seega on keeltevahelisel sarnasusel põhinev analoogia keeleomandamise keskne psühholingvistiline mõiste, mistõttu pole ei teise keele õpetamisel ega ka selle omandamise protsessi uurimisel mõttekas lähendada emakeele ja õpitava keele erinevustest (kontrastiivne keelekäsitlus, interferentsiteooria), vaid nende sarnasusest. Inimese keelevõime, suutlikkus aru saada, mõistestada ja teadmisi talletada tugineb assotsiatsioonidele ning analoogiapõhistele üldistustele ning on vahetult seotud nende protsesside aluseks oleva teadmuse universaalse iseorganiseerimisvõimega.

### 3.2. Kontseptuaalne semantika

Kontseptuaalse semantika raames käsitletud teoreetiliste lähenemiste taustal on analoogia mõistel oluline tähendus, eriti silmas pidades esimese ja teise keele omandamist.

Esiteks seetõttu, et analoogia on arusaamise, teadmuse kujunemise ja edastamise efektiivne vahend. Sõnade ja väljendite mõistestamisel tugineb inimene suuresti oma kogemusele ning sellest tulenevalt metafoorsele, metonüümsele vm laadi analoogiapõhisele mõtlemisviisile, mis motiveerib meie keelekasutust oluliselt suuremal määral, kui arvatud. See ilmneb sõnade, vormide ja süntaksistruktuuride kasutamises ülekantud tähenduses (fraseologismid, registrierinevustega seotud vormivariatiivsus jne) või neile mitteomases funktsioonis (nt *praesens historicum* või verbi 2. pöörde üldistatud kasutamine kelle tahes suhtes, kes situatsiooni kogevad: *See kordumatu igaviku tunne, kui istud niisugustel sügisõhtutel kodusoojas ja kuulad raugeva tormi müha!*). Neid nähtusi ei saa pidada anomaaliaks: tegu on keelele omase iseorganiseerumisvõime ilmnemisega, tänu millele on omavahel tasakaalus keele säilimine ja pidev tsüklikiline uuenedmine.

Teiseks on analoogia mõiste oluline metodoloogiliselt, kuna kujundliku analoogiapõhise mõtlemisviisi eesmärgipärane ja teadlik rakendamine võimaldab kasutada analoogiat nii universaalse uurimismeetodina (vt Itkonen 2005) kui ka keelekirjelduste alusena (vt ka Eslon 2006). Nii näiteks osutatakse kognitiivse metafooriteooria raames keeleüksuste semantikas kajastuvale kogemusele metafoori allikvaldkondade kaudu. Erinevate metakeelsete mõistemetafooride abil püütakse välja tuua analoogia alusel tekkinud tähendusülekannete suunad ja olemus (vt Lakoff & Johnson 1999: 161–164).

Ungari kognitivist Zoltán Kövecses on pidanud kontseptuaalsetele analoogiapõhiste mõisteseoste tuginevat fraseologismide korraldamist sõnavara õppimise ja õpetamise oluliseks osaks (Kövecses 2000). Kövecsesi seisukohta saab mõista ka avaramalt: nii esimese keele omandaja kui ka teise keele õppija tuginevad mõisteseoste ühtmoodi, kui

nad idiomaatilisi üksusi seostavad, nende olemust lahti mõtestavad või vastavalt suhtlusvajadustele kasutavad. Mõisted ja nendevahelised seosed moodustavad universaalse semantilise võrgustiku ehk semantilise välja, millel erinevates keeltes võib olla kas sarnane või keelemeelele omane sõnaline ja grammatiline väljendus, kuid mõisteline alus, millest lähtudes sõna- või vormivalikuid tehakse, on üldplaanis sama. Tänu ühtsele universaalsele alusmõistestikule saab hinnata, kuidas emakeelekõneleja idiomaatikat tegelikult kasutab ja kuidas teise keele õppija seda oskust oma mentaalse leksikoni alusel järk-järgult omandab.

Sel teadmisel on lisaks muule ka oluline pedagoogiline tähendus. Näiteks I–IV klassi emakeele õppematerjalide analüüs (Puhtvend 2009) tõi esile eesti õpikute ja töövihikute materjalivaliku juhuslikkuse ning sisulise põhjendamatus. Eesti fraseologisme, ülekantud tähenduses sõnu ja fraase õpetatakse meil tegelikult üksikkeelendist või -juhtumist lähtuvalt, mitte mõisteseoste alusel. Hoopis tähelepanuta on jäänud nii keeleomane kui ka üldinimlik kujundisüsteem ja selle tekkemehhanismid. Idiomaatiliste üksuste mõistelise motivatsiooni eksplitseerimine juhiks õppijat nende omandamisel rakendama nii oma maailma- kui ka oma emakeelest tulenevat keelelist teadmust, mis aitaksid õpitavat seletada, kujundada sarnasusel ja analoogial põhinevaid seoseid, sellele tuginedes väljendeid meelde jätta ja õpitava keele kujundisüsteemi enda jaoks uuesti üles ehitada (vt eespool MacWhinney universaalne mudel). Kui osutada õppija jaoks tundmatu idiomaatilise üksuse aluseks olevatele, kuid emakeelest tuttavatele kujunditele, aitaks see võõra väljendi tähendust paremini mõista ja meelde jätta.

Kontseptuaalse semantika rakenduslikke väljundeid on WordNet-tüüpi sõnastikud ehk tesaurused, arvutileksikograafia, ontoloogiad, intelligentset arvutisüsteemid (nagu eespool mainitud dialoogsüsteem DST) jne, kus tegeldakse ühel või teisel moel inimese teadmiste modelleerimisega.

### **3.3. Keele töödeldavusteooria, ühtse võistlusmudeli ja kontseptuaalse semantika integreerimine kasutuspõhises keeleuurimises**

Keele töödeldavusteooria, ühtse võistlusmudeli, keeltevahelise mõju mehhanismi ja kontseptuaalse semantika integreerimine toetub kasutus- põhisuse printsiibile: mis tahes keelt õpitakse ja rakendatakse suhtlus- vajadustest lähtuvalt ning keelekasutuse kaudu.

Ühelt poolt võimaldab see vaadelda ja analüüsida keelesüsteemi seoste võrgustikuna, mis moodustub semantiliste kontseptide vahel välja piires ning väljade vahel. Semantilistele kontseptidele ja nende vahelistele seostele vastavad keelekasutuses kindlate leksikaalsete ja morfosüntaktiliste nähtuste kogumid, kus ükski element ega struktuur ei eksisteeri teistest sõltumata. Ka teise keele õppimist võib vaadelda ja analüüsida samalaadse semantiliste seoste võrgustiku alusel, ainult selle vahega, et lähtudes emakeelest ehitab õppija võrgustiku uuesti üles. Nii või teisiti on siin tegu universaalse semantiliste seoste võrgustikuga, millele keelekasutuses vastavad teatud regulaarsusega ilmnevad semantilised ja morfosüntaktilised terviküksused ehk konstruktsioonid. Just konstruktsiooni (mitte lekseemi, kategooriat, paradigmat, keelestruktuuri või reeglit) on peetud psühholoogiliseks ja ka teoreetiliseks entiteediks (vt Tomasello 2006: 258; Wilson 2003: 83; Leino 2007; Pajusalu & Tragel 2007). Lapsed, kes on omandanud muutmissüsteemi kindlate konstruktsioonide ja leksikaalsete üksuste kaudu, hakkavad edaspidi iseseisvalt konstrueerima grammatilisi mooduleid (Dressler 2004: 9), valides selleks sobivaid keeleüksusi, kusjuures erinevatel keeleomandamisastmetel juhivad nende valikuid erinevad vajadused (vt Bittner jt 2003: xviii). Sama on täheldatud ka teise keele omandamisel (vt Kaivapalu 2005).

Seetõttu on keeleomandamise uurimises olulised konstruktivistide seisukohad (vt Chang jt 2002; Boas 2002; Östman & Fried 2005; Dressler & Karpf 1995; Karpf 1991 jt). Nemad määratlevad konstruktsiooni kui keelelist üksust, mille tähendust, süntaktilist struktuuri ja leksikaalset täitumust ei vaadelda eraldi, vaid kontiinumina (Croft 2001: 16–17).

Konstruksiooni lingvistiline kirjeldus sisaldab niisugust morfosüntakti- list kirjet, millest ilmneb konstruksiooni konventsionaalne tähendus ehk kinnistatus teatud semantilise kontseptiga. Keeles on “keeleväljendeid, mille struktuur ei taandu neid moodustavate leksikaalsete üksuste kombinatoorsetele omadustele ja üldistele grammatilistele seaduspäradele” ning “idiomaatilist --- tähendust ei pruugi kanda mitte ainult --- kinnis- tunud sõnaühendid, vaid ka produktiivselt moodustatud väljendid” (vt Sahkai 2008: 171–172) ehk morfosüntaktilised konstruksioonid. Näiteks eesti adessiivse viisi- ja põhjsumääruse konstruksioonid nagu *Rahva tungival soovil loosib hiphop.ee ühe plaadi ka oma küllastajate vahel välja; Eesti Kirjameeste Seltsi otsusel mindi 1872 üle uue kirjaviisi kasutamisele; Magistrikursuslased õpivad telesilla abil* jne (vt Sahkai 2006) või vahen- datud väite edastamiseks sageli kasutatavad konstruksioonid (vt Sepper 2006: 25 jj), milles esinevad vaid valitud vormid ja sõnavara: *Tema sõnul on küll Venemaal suur pidudeaeg ning tööle tullakse alles 10. jaanuaril, kuid Eesti saatkond mõtleb juba tööle; Sampo Panga hinnangul otsus- tabki just kinnisvaraturu edasine toimimine, kas Eesti majandust ootab pehme või kõva maandumine; Simoni kinnitusel jäi tookordne istung ära tema haigestumise tõttu* jne (vt Kostenko 2010: 23 jj). Sedalaadi piirangud on keelekasutuses tunduvalt suuremad kui fraseoloogilised kitsendused. Arvo Krikmann on kinnitanud, et tema soome ja eesti folkloori- ning keelearhiivide empiirilisest uurimisest ammutatud kogemus näitab, et “põhiosa folkloorist --- on variaabel folkloor ja et klišeeline folkloor on pigem erand, mis kinnitab reeglit. --- Me kohtame folkloorses reper- tuaaris püsikujundeid, mis võivad istutada end väga erineva süntaktilise struktuuriga süntagmadesse ja lausetesse, üldse väga erinevaise kon- tekstidesse. Teisalt nähtub süntaktilisi püsimalle, mille “vabu valentse” on võimalik täita kümnete, sadade või tuhandete erinevate figuratiivsete vm leksikaalsete elementidega”. Valdava arusaama klišeelisusest loovad sõnaraamatud ning ainult keelecorpuse alusel saab näha keelekasutuse tegelikku varieerumist (Krikmann 2004: 106).

Arusaama sagedastest ning keelekasutusele iseloomulikest konst- ruksioonidest on suuresti aidanud kujundada kaasaegsete arvuti-

tehnoloogiate ja statistiliste meetodite rakendamine mahuka korpusainese keele- ja diskursusanalüüsis (vt Barnbrook 1998; Biber jt 1998; Sinclair 2004; Scott & Tribble 2006). See on loonud uut laadi korpusandmetest lähtuva keelekirjelduse, mis on toonud uusi võimalusi sõnavormide distributsiooni, kontekstitundlikkuse, sünonüümsuse, keelekasutaja eelistuste, sageli kasutatavate vabade sõnaühendite, idiomaatiliste üksuste ja kollokatsioonide leidmiseks korpusainesest ning nende lingvistiliseks interpreteerimiseks. Olgugi et neid keeleüksusi on nimetatud erinevalt<sup>13</sup>, võib nende kasutamises erinevates diskursusetüüpides, keelevariantides (nt esimene keel, teine keel, murdekeel), allkeeltes (nt ilukirjandus, publitsistika, teaduskeel, kõnekeel) näha esinemissagedusele tuginevaid mustreid (vt näiteks Sinclair 2004; Stubbs 1996; Altenberg & Granger 2001; Jantunen 2004; Masini 2005), mille põhjal saab võrrelda erinevate keelekasutusvariantide ühisjooni, spetsiifikat ja lingvistilist varieerumist. Näiteks on tehtud katse välja tuua EVKK-s ning EKI korpuses eelistatud keelekasutusmustrid ja neis esinenud käändevormid. Siinkohal tuuakse näiteid konstruktsioonidest, kus sõna *inimene* on sageli ainsuse nominatiivis (vt Eslon 2009: 40–42):

#### EVKK

- 1) loogiline implikatsioon: *kui inimene on* (variandid: *kui inimene tahab, kogeb, elab, suhtleb*) ~ *et inimene peab* ~ ja *kui* (*et kui, sest kui*) *inimene*
- 2) modaalkonstruktsioon: *inimene peab olema* (*maksma, teadma*) ~ *inimene võib töötada; haritud (iga) inimene peab* (*saab, võib, tahab*)
- 3) eitust sisaldav (modaal)konstruktsioon: *inimene ei saa* (*ei ole*)

<sup>13</sup> Nt sageduse alusel välja toodud süntaktilisteks ja semantilisteks terviküksusteks (vt Nesselhauf 2005: 21), mitmesõnalisteks väljenditeks (vt Granger 2005; Masini 2005), erinevat laadi idiomaatilisteks üksusteks ja püsiühenditeks (Muischnek 2006) jne.

EKI korpus

- 1) substantiivkonstruktsioon: *inimene ja seadus ~ inimene ja loodus*
- 2) loogiline implikatsioon: *kui inimene on ~ et inimene on*
- 3) eitust sisaldav (modaal)konstruktsioon: *inimene ei ole (ei saa)*

Näide sisaldab esimese ja teise keele tekstiloomes sagedasi konstruktsioone, mis on omased eesti keelt emakeelena kõnelejale ja eesti keele kui teise keele õppijale. Selgelt tuleb esile vahe kasutuseelistustes (õppijakeeles on eelistatud modaal- ja kirjakeeles substantiivkonstruktsioone) ning samalaadsete morfosüntaktiliste konstruktsioonide leksikaalgrammatilises varieerumises (vt loogiline implikatsioon). Nimisõnade ainuse nominatiivi kasutuse põhjalikum analüüs peaks välja tooma, kas see nähtus on kahe keelevariandi jaoks tüüpiline või mitte. Järelikult on sagedusel rajanev kasutuspõhisus üks olulisi lähtekohti mitte ainult erinevate (keeleomandamise) teooriate ja uurimisvaldkondade integreerimisel, vaid ka esimese ja teise keele oskuse järk-järgulise arenemise ning tasemeoskuste lingvistilise sisu kindlaks tegemisel.

#### 4. Kokkuvõttev arutelu

Kavandatava esimese ja teise keele võrdleva uurimise teoreetiliste seisukohtade avamine tõi esile sellised olulised märksõnad nagu kasutuspõhisus, analoogia, universaalsus, integreeritus, mitmemõõtmelisus. Nii Pienemanni ja MacWhinney keeleomandamise teooriad kui ka Ringbomi jt keeltevahelise mõju psühholingvistiliste mehhanismide uurimused rajanevad arusaamal inimese teadmuse talletamisest pikaajalises mälus teadmismustritena, mis eksplitseeruvad keelekasutuses, ilmnevad keelelise väljenduslaadi valikutest (keelemustritena) ja keelekasutaja eelistuste kaudu, kui semantiline kontsept seotakse sellele vastava predikaadi argumentidestruktuuriga. See on keele produtseerimise universaalne alus, mille põhjal saab uurida ning kirjeldada mistahes keele omandamist ja kasutamist mistahes etappidel ja tasemetel.

Keele omandamise seisukohalt on oluline mentaalne leksikon, milles teadmus on talletatud semantiliste kontseptide vaheliste mõisteseoste alusel. See universaalne semantiliste seoste võrgustik määrab produtseerimis- ja arusaamisprotsessi keelelise sisu: lause argumendistruktuuri, intonatsiooni ja hääldusviisi, aja, aspekti, isiku, arvu, objekti- ja subjektiivkäände valiku jm, kusjuures kesksed on niisugused keelekategoriad nagu verb, substantiiv, adjektiiv. Verbikesksete konstruktsioonide kasutuses on otsustav osa verbi funktsionaalsel potentsiaalil, mida mõjutab fraasisisene ja lause sõnajärg. Siit johtub ka nende konstruktsioonide omandamise keerukus ja probleemsus, sest verbisüntaksis lõikuvad morfoloogia, süntaks, semantika ja pragmaatika.

Keeleüksusi kombineeritakse valivalt. Emakeeles tingib ühe või teise grammatilise vormi eelistamine teatud morfosüntaktilistes kompleksides teiste teatud vormide ilmnemise (keeleelementide modulaarne organiseeritus), mis leiab kajastust ennekõike interpretatiivsetes funktsionaalsemantilistes kategooriates nagu aeg, aspekt, modaalsus. Aga teise keele õppimisel tuleb see, kuidas keeleüksusi valida ja järjestada, meelde jätta. See teadmus talletatakse lühimälus assotsiatiivse sidususe alusel oma mentaalse leksikoniga. Kuna lühimälu aluseks olevad kontseptid on üldised, siis võivad assotsiatsiooniseosed oma mentaalse leksikoniga olla (ligi)lähedased või ekslikud. Paradoks on selles, et teise keele õppimine ja omandamine rajaneb just nendel universaalsetel kontseptidel ja hägusatel assotsiatsiooniseostel. Analoogiale tugineva positiivse keeleülekande rakendumisel annab see keeleõppes häid tulemusi.

Inimese keelevõime, suutlikkus aru saada, mõistestada ja teadmisi talletada tugineb assotsiatsioonidele ning analoogiapõhistele üldistustele ning on vahetult seotud nende protsesside aluseks oleva teadmuse universaalse iseorganiseerumisvõimega. Siit tuleneb esimese ja teise keele omandamise loogika ning produtseerimisprotsessi olemus, mille ühis- ja erijooni saab välja tuua dialoogsüsteemi DST abil. Analüüsi tulemusel saadakse ettekujutus ühelt poolt esimese ja teise keele oskuse arengust, selle protsessi ühis- ja erijoonest ning teisalt erinevate suhtluseesmärkide täitmiseks vajalike keeleelementide modulaarsest organiseeritusest.

Kuna analoogia on arusaamise, teadmuse kujunemise ja edastamise efektiivne vahend, siis võib seda mõistet pidada ka oluliseks metodoloogia seisukohalt: kujundliku analoogiapõhise mõtlemisviisi eesmärgipärane ja teadlik rakendamine võimaldab kasutada analoogiat nii universaalse uurimismeetodina (sh keele omandamise ja kasutamisega seotud protsesside modelleerimisel) kui ka keelekirjelduste alusena. Seda toetavad konstruktivistide arusaamad keeleüksuste eelistamisest, vormide järk-järgulisest ilmnemisest kindlat liiki konstruktsioonides jm, mida aitab esile tuua korpuslingvistiline keeleanalüüs. Keeleoskustasemete lingvistilise sisu ja taseme määramise teeb täpsemaks see, kui eelnevalt on vähemalt kümne<sup>14</sup> eksperdi töö tulemusena tehtud kindlaks analüüsitava tekstide tase (nt A1, A2, B1, B2 jne) ning enamus eksperte on samal arvamusel. Kui seejärel koondada iga taseme tekstid vastava taseme korpuseks, analüüsida neid alamkorpusi statistikapõhiste standardprogrammide ja eesti keele tarkvaraga ning töödelda saadud tulemusi statistiliselt, siis peaksid esile tulema tasemele iseloomulikud keelemoodulid ehk idiomaatiliste üksuste ning morfosüntaktiliste konstruktsioonide kogumid. Selle põhjal saab esimese ja teise keele omandamist võrrelda tasemeti, tuues keelekasutuse alusel välja kokkulangevused, analoogsed jooned ja eripära.

## Tänu sõnad

Artikli valmimist on toetanud riikliku programmi “Eesti keele keeletehnoloogiline tugi (2006–2010)” projektid “VAKO: Eesti vahekeele korpuse keeletarkvara ja keeletehnoloogilise ressursi arendamine (2008–2010)” ja “Eesti fraseologismide elektroonilise alussõnastiku loomine (2008–2010)”; ETF-i grant 8222 “Ülekantud tähenduses fraasid eesti keele korpustes (2010–2013)”; ETF-i grant 8240 “Lähteelele mõju ja teise keele omandamine (korpuspõhine uurimus)”; rahvusvaheline

<sup>14</sup> Tasemeoskuste määramisest ning selle sidumisest keeletestide standardimisega saab ettekujutuse 2003. aasta Euroopa Nõukogu dokumendist “Keeleeksamite sidumine Euroopa keeleõppe raamdokumendiga: õppimine, õpetamine ja hindamine”, mille ingliskeelset versiooni on täiendatud 2010. aastal. Eestikeelse tööversiooni kasutamise võimaluse eest tänusõnad REKK-i keeleosakonnale.

projekt COST A33 “Crosslinguistically Robust Stages of Language Development”; riikliku programmi “Eesti keel ja kultuurimälu (2009–2013)” projekt “REKKi käsikirjaliste materjalide digiteerimine, Eesti vahekeele korpuse alamkorpuste loomine ja korpuse kasutusvõimaluste populariseerimine (2009–2013)”.

## Kirjandus

- Abney, Steven 1996. Statistical methods and linguistics. – Judith Klavans, Philip Resnik (Eds.). *The Balancing Act: Combining Symbolic and Statistical Approaches to Language*. Cambridge, MA: The MIT Press, 1–23. <http://www.vinartus.net/spa/95c.pdf> (11.09.2008).
- Alderson, Charles J. 2006. *Diagnosing Foreign Language Proficiency. The Interface between Learning and Assessment*. London, New York: Continuum International Publishing Group.
- Alderson, Charles J. 2007. The CEFR and the need for more research. – *The Modern Language Journal* 91 (4), 659–663. doi:[10.1111/j.1540-4781.2007.00627\\_4.x](https://doi.org/10.1111/j.1540-4781.2007.00627_4.x)
- Alderson, Charles J., Ari Huhta 2005. The development of a suite of computer-based diagnostic tests based on the Common European Framework. – *Language Testing* 22, 301–320. doi:[10.1191/0265532205lt310oa](https://doi.org/10.1191/0265532205lt310oa)
- Alp, Pilvi 2004. Eesti keele algtaseme testi arendamise võimalusi. Magistritöö. Käsikiri Tallinna Ülikooli eesti keele ja kultuuri instituudis.
- Altenberg, Bengt, Sylviane Granger 2001. The grammatical and lexical patterning of make in native and non-native student writing. – *Applied Linguistics* 22 (2), 173–194. doi:[10.1093/applin/22.2.173](https://doi.org/10.1093/applin/22.2.173)
- Argus, Reili 2008a. Eesti keele muutemorfoloogia omandamine. Tallinna Ülikooli humanitaarteaduste dissertatsioonid 19. Tallinn: Tallinna Ülikooli Kirjastus. <http://www.tlu.ee/files/news/1273/Reili8d970cc1dd8a7713e61-de7b43c974d71.pdf> (26.08.2008).
- Argus, Reili 2008b. Pühholingvistiline katse eesti keele objekti käändevahelduse omandamise uurimise meetodina. – *Emakeele Seltsi aastaraamat* 54, 22–43.
- Argus, Reili 2009. Eksperimentaalse metoodika kasutamisest eesti keele omandamise alastes uuringutes. – Pille Eslon, Katre Õim (Toim.). *Korpusuuringute metodoloogia ja märgendamise probleemid*. Tallinna Ülikooli Eesti Keele ja Kultuuri Instituudi toimetised 11. Tallinn: Tallinna Ülikooli Kirjastus, 7–29.

- Asser, Hiie, Heiki-Jaan Kaalep, Siret Linna, Jaan Mikk, Kadri Muischnek, Merje Songe, Heli Uibo 2004. Õpikute keerukuse analüüs arvutitel. – Helle Metslang, Maria-Maren Sepper, Jane Lepasaar (Toim.). Toimiv keel II: Tööd rakenduslingvistika alalt. Tallinn: Tallinna Pedagoogikaülikooli Kirjastus, 72–84.
- Atkins, Sue, Beth Levin 1995. Building on a corpus: A linguistic and lexicographical look at some near-synonyms. – *International Journal of Lexicography* 8 (2), 85–114. doi:[10.1093/ijl/8.2.85](https://doi.org/10.1093/ijl/8.2.85)
- Baker, Paul 2006. *Using Corpora in Discourse Analysis*. London: Continuum.
- Barnbrook, Geoff 1998. *Language and Computers: A Practical Introduction to the Computer Analysis of Language*. Edinburgh: Edinburgh University.
- Bayley, Robert, Dennis R. Preston (Eds.) 1996. *Second Language Acquisition and Linguistic Variation. Studies in Bilingualism 10*. Amsterdam: John Benjamins.
- Behrens, Heike (Ed.) 2008. *Corpora in Language Acquisition Research: History, Methods, Perspectives. Trends in Language Acquisition Research 6*. Amsterdam, Philadelphia: John Benjamins.
- Biber, Douglas 1993. Representativeness in corpus design. – *Literary and Linguistic Computing* 8 (4), 243–257. doi:[10.1093/lc/8.4.243](https://doi.org/10.1093/lc/8.4.243); 2nd ed. in: Thierry Fontenelle (Ed.). *Practical Lexicography: A Reader*. Oxford: Oxford University Press, 63–88.
- Biber, Douglas 2004. Conversation text types: A multi-dimensional analysis. – JADT: 7es Journées internationales d'Analyse statistique des Données Textuelles. Louvain, 15–34. [http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2004/pdf/JADT\\_000.pdf](http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2004/pdf/JADT_000.pdf) (9.05.2009).
- Biber, Douglas, Susan Conrad, Randi Reppen 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Bittner, Dagmar, Wolfgang U. Dressler, Marianne Kilani-Schoch 2003. Introduction. – Peter Jordens (Ed.). *Development of Verb Inflection in First Language Acquisition. A Cross Linguistic Perspective. Studies on Language Acquisition 21*. Berlin, New York: Mouton de Gruyter, 7–26.
- Boas, Hans C. 2002. On constructional polysemy and verbal polysemy in construction grammar. – Vida Samiian (Ed.). *Proceedings of the 2000 Western Conference on Linguistics*, vol. 12, 126–139.
- Chambers, John M. 2007. *Software for Data Analysis: Programming with R*. New York: Springer.
- Chang, Nancy, Jerome Feldman, Robert Porzel, Keith Sanders 2002. Scaling Cognitive Linguistics: Formalisms for Language Understanding. <http://www.icsi.berkeley.edu/NTL/papers/scaling.pdf> (10.05.2010).

- Cobb, Tom 2003. Analyzing late interlanguage with learner corpora: Québec replications of the three European studies. – *The Canadian Modern Language Review* 59 (3), 393–423. doi:[10.3138/cmlr.59.3.393](https://doi.org/10.3138/cmlr.59.3.393)
- Croft, William 2001. *Radical Construction Grammar*. Oxford: Oxford University Press. doi:[10.1093/acprof:oso/9780198299554.001.0001](https://doi.org/10.1093/acprof:oso/9780198299554.001.0001)
- Dressler, Wolfgang U. 2004. Morphological typology and first language acquisition: Some mutual challenges. – Geert Booij, Emiliano Guevara, Angela Ralli, Salvatore SgROI, Sergio Scalise (Eds.). *Morphology and Linguistic Typology. On-line Proceedings of the Fourth Mediterranean Morphology Meeting, Catania, September 20–23, 2003*, 8–20. <http://mmm.lingue.unibo.it/proc-mmm4.php> (31.10.2007).
- Dressler, Wolfgang U., Annemarie Karpf 1995. The theoretical relevance of pre- and protomorphology in language acquisition. – Geert Booij, Jaap van Marle (Eds.). *Mechanisms of Morphological Change. Yearbook of Morphology 1994*. Dordrecht: Kluwer, 99–122.
- Eesti keele arendamise strateegia 2004–2010. Tartu 2004. <http://www.hm.ee/index.php?popup=download&id=4149> (18.09.2008).
- Eslon, Pille 2006. Analoogiast keelte kõrvutamisel. – *Keel ja Kirjandus* 1, 15–24.
- Eslon, Pille 2009. Eestikeelses tekstiloomes eelistatud konstruktsioonid ja käändevormid. – Pille Eslon, Katre Õim (Toim.). *Korpusuuringute metodoloogia ja märgendamise probleemid*. Tallinna Ülikooli eesti filoloogia osakonna toimetised 11. Tallinn: Tallinna Ülikooli Kirjastus, 30–53.
- Eslon, Pille, Erika Matsak 2009. Eesti keele kasutusvariandid: korpusest tulenev ja korpuspõhine võrdlev analüüs. – *Eesti Rakenduslingvistika Ühingu aastaraamat* 5, 79–110. doi:[10.5128/ERYa5.06](https://doi.org/10.5128/ERYa5.06)
- Genoz, Jasone, Britta Hufeisen, Ulrike Jessner 2001. *Gross-linguistic Influence in Third Language Acquisition: Psycholinguistic Perspectives*. Clevedon: Multilingual Matters LTD.
- Granger, Sylviane 1997. On identifying the syntactic and discourse features of participle clauses in academic English: Native and non-native writers compared. – Jan Aarts, Inge de Mönnink, Herman Wekker (Eds.). *Studies in English Language and Teaching*. Amsterdam: Rodopi, 185–198.
- Granger, Sylviane 1998. Prefabricated patterns in advanced EFL writing: Collocations and formulae. – Anthony Paul Cowie (Ed.). *Phraseology: Theory, Analysis, and Applications*. Oxford: Oxford University Press, 145–243.
- Granger, Sylviane 2003. Error-tagged learner corpora and CALL: A promising synergy. – *CALICO Journal* 20 (3), 465–480.

- Granger, Sylviane 2004. Computer learner corpus research: Current status and future prospects. – Ulla Connor, Thomas A. Upton (Eds.). *Applied Corpus Linguistics: A Multidimensional Perspective*. Amsterdam, Atlanta: Rodopi, 123–145. <http://www.fltr.ucl.ac.be/fltr/germ/etan/cecl/Downloads/India-apolis%20status%20&%20prospects.pdf> (25.06.2008).
- Granger, Sylviane 2005. Pushing back the limits of phraseology: How far can we go? – Christelle Cosme, Céline Gouverneur, Fanny Meunier, Magali Paquot (Eds.). *Proceedings of the phraseology 2005 conference*, Louvain-la-Neuve, 13–15 October 2005. Louvain-la-Neuve: Presses universitaires de Louvain, 165–168.
- Habicht, Külli 2006. Meie muutuv kirjakeel. – *Oma Keel* 1, 12–19.
- Hanston, Susan 2002. Pattern grammar, language teaching, and linguistic variation: Applications of a corpus-driven grammar. – Randi Reppen, Susan Fitzmaurice, Douglas Biber (Eds.). *Using Corpora to Explore Linguistic Variation*. *Studies in Corpus Linguistics* 9. Amsterdam, Philadelphia: John Benjamins, 167–186.
- Hausenberg, Anu-Reet 2003. Eesti keele teise keelena uurimisest: mida ja milleks. – Margit Langemets, Heete Sakhai, Maria-Maren Sepper (Toim.). *Toimiv keel I. Töid rakenduslingvistika alalt*. Eesti Keele Instituudi toimetised 12. Tallinn: Eesti Keele Sihtasutus, 53–64.
- Hausenberg, Anu-Reet, Marju Ilves, Annekatrin Kaivapalu, Krista Kerge, Katrin Kern, Mare Kitsnik, Ingrid Krall, Karin Rummo, Tiina Rüütmaa 2008. *Iseseisev keelekasutaja: B1- ja B2-taseme eesti keele oskus*. Tallinn: Riiklik Eksami- ja Kvalifikatsioonikeskus.
- Hoey, Michael, Michaela Mahlberg, Michael Stubbs, Wolfgang Teubert (Eds.) 2007. *Text, Discourse, and Corpora: Theory and Analysis*. With an Introduction by John Sinclair. London, New York: Continuum International Publishing Group. <http://www.icsi.berkeley.edu/NTL/papers/scaling.pdf> (28.08.2008).
- Hulstijn, Jan H. 2007. The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency. – *The Modern Language Journal* 91, 663–667. doi:[10.1111/j.1540-4781.2007.00627\\_5.x](https://doi.org/10.1111/j.1540-4781.2007.00627_5.x)
- Ilves, Marju 2008. *Algaja keelekasutaja. A2-taseme eesti keele oskus*. Tallinn: Eesti Keele Sihtasutus.
- Ilves, Marju 2010. *Läbimurre. A1-taseme eesti keele oskus*. Tallinn: Eesti Keele Sihtasutus.
- Itkonen, Esa 2005. *Analogy as Structure and Process*. *Approaches in Linguistics, Cognitive Psychology and Philosophy of Science*. *Human Cognitive Processing* 14. Amsterdam, Philadelphia: John Benjamins.

- Jantunen, Jarmo Harri 2004. Synonymia ja käännösuomi. Korpusnäkökulma samamerkityksisyyden kontekstuaalisuuteen ja käännöskielen leksikaaliin erityispiirteisiin. Joensuun yliopiston humanistisia julkaisuja 35. Joensuu: Joensuun yliopisto.
- Jantunen, Jarmo Harri 2008. Haasteita oppijankielen korpusanalüysille: oppijankielen universaalit. – Pille Eslon (Toim.). Õppijakeele analüüs: võimalused, probleemid, vajadused. Tallinna Ülikooli eesti filoloogia osakonna toimetised 10. Tallinn: Tallinna Ülikooli Kirjastus, 67–92.
- Jarvis, Scott, Terrence Odlin 2000. Morphological type, spatial reference, and language transfer. – *Studies on Second Language Acquisition* 22, 535–556. doi:[10.1017/S0272263100004034](https://doi.org/10.1017/S0272263100004034)
- Jürine, Anni 2007. Eesti kaassõnaühendite norminguvastases kokkukirjutamisest grammatiseerumisteooria taustal. Bakalaureusetöö. Käsikiri Tartu Ülikooli eesti keele osakonnas.
- Jüssi, Elise 2005. Laadivahelduse taandumine eesti kirjakeeles õpilaste keelekasutuse näitel. Bakalaureusetöö. Käsikiri Tartu Ülikooli eesti keele osakonnas.
- Kaalep, Heiki-Jaan, Kadri Muischnek 2002. Eesti kirjakeele sagedussõnastik. Tartu: Tartu Ülikooli Kirjastus.
- Kaivapalu, Annekatrin 2005. Lähdekeeli kielennõppimise apuna. *Jyväskylä Studies in Humanities* 44. Jyväskylä: University of Jyväskylä.
- Kaivapalu, Annekatrin 2009. Õppijakeele korpusanalüüsi täiendavatest meetoditest. – Pille Eslon, Katre Õim (Toim.). Korpusuuringute metodoloogia ja märgendamise probleemid. Tallinna Ülikooli eesti keele ja kultuuri instituudi toimetised 11. Tallinn: Tallinna Ülikooli Kirjastus, 72–98.
- Kaivapalu, Annekatrin, Maisa Martin 2007. Morphology in transition: The plural inflection of Finnish nouns by Estonian and Russian learners. – *Acta Linguistica Hungarica* 54 (2), 129–156. doi:[10.1556/ALing.54.2007.2.2](https://doi.org/10.1556/ALing.54.2007.2.2)
- Karpf, Annemarie 1991. Universal grammar needs organization. – *Folia Linguistica* XXV/3–4, 339–360. doi:[10.1515/flin.1991.25.3-4.339](https://doi.org/10.1515/flin.1991.25.3-4.339)
- Kerge, Krista 2008. Vilunud keelekasutaja. C1-taseme eesti keele oskus. Tallinn: Eesti Keele Sihtasutus.
- Keßler, Jörg-U. 2009. Arbeitstagung Englischlernern als Kontinuum, Düsseldorf, 26.5.2009. [http://www.medienberatung.nrw.de/fachthema/dokumentationen/ws\\_kessler\\_fruehbeginn.pdf](http://www.medienberatung.nrw.de/fachthema/dokumentationen/ws_kessler_fruehbeginn.pdf) (1.11.2009).
- Kilgarrieff, Adam 2001. Comparing corpora. – *International Journal of Corpus Linguistics* 6 (1), 97–133. doi:[10.1075/ijcl.6.1.05kil](https://doi.org/10.1075/ijcl.6.1.05kil). <http://www.kilgarrieff.co.uk/Publications/2001-K-CompCorpIJCL.pdf> (21.12.2009).

- Kilgarriff, Adam 2003. What computers can and cannot do for lexicography, or, us precision, them recall. Keynote lecture. Proc. ASIALEX. Tokyo, August. <http://www.kilgarriff.co.uk/publications.htm> (24.11.2009).
- Kitsnik, Mare 2006. Keelekorpused ja võõrkeeleõpe. – Eesti Rakenduslingvistika Ühingu aastaraamat 2, 93–107.
- Kitsnik, Mare 2007. Öppijakeele uurimise ja arendamise võimalusi eesti vahekeele korpuse põhjal (eituse väljendamise näitel). Magistritöö. Käsikiri Tallinna Ülikooli eesti keele ja kultuuri instituudis.
- Kostenko, Anne 2010. Postimees-online ajakirjandustekstide tõlkeanalüüs: vahendatud väite tõlkimisvõimalusi. Tallinn. Käsikiri Tallinna Ülikooli romaani ja germaani keelte instituudis.
- Krikmann, Arvo 2004. “Sai hea obaduse vastu obadust”: Löömist ja peksmist märkivad väljendid eesti keeles. – Reetor 3. Tartu: Eesti Kirjandusmuuseum Folkloristika osakond, Eesti Kultuuriloo ja Folkloristika Keskus.
- Kövecses, Zoltán 2000. A Cognitive Linguistic View of Learning Idioms in an FLT Context. Essen: LAUD Linguistic Agency, University-GH Essen.
- Lakoff, George, Mark Johnson 1999. *Metaphors We Live By*. Chicago, London: The University of Chicago Press.
- Laufer, Batia 2000. Avoidance of idioms in a second language: The effect of L1-L2 degree of similarity. – *Studia Linguistica* 54 (2), 186–196. doi:[10.1111/1467-9582.00059](https://doi.org/10.1111/1467-9582.00059)
- Leech, Geoffrey 2004. Recent grammatical change in English: data, description, theory. – Karin Aijmer, Bengt Altenberg (Eds.). *Advances in Corpus Linguistics: Papers from the 23rd international conference on English language research and computerized corpora (ICAME 23)* Göteborg, 22–26 May 2002. Amsterdam: Rodopi, 61–81.
- Leech, Geoffrey, Paul Rayson, Andrew Wilson 2001. *Word Frequencies in Written and Spoken English: Based on the British National Corpus*. London: Longman. <http://ucrel.lancs.ac.uk/bncfreq/> (25.06.2008).
- Leino, Jaakko (Ed.) 2007. *Constructional Reorganization. Constructional Approaches to Language 5*. Philadelphia: John Benjamins.
- Leistyna, Pepi, Charles F. Meyer (Eds.) 2003. *Corpus Analysis: Language Structure and Language Use*. Amsterdam: Rodopi.
- Lepajõe, Kersti 2006. Õpilaste registri- ja stiilitajust riigieksamikirjandite põhjal. – Ellen Niit (Toim.). *Keele ehe*. Tartu Ülikooli eesti keele õppetooli toimetised 30. Tartu: Tartu Ülikool, 141–153.
- Lindquist, Hans, Cristian Mair (Eds.) 2004. *Corpus Approaches to Grammaticalization in English. Studies in Corpus Linguistics 13*. Amsterdam, Philadelphia: John Benjamins.

- MacWhinney, Brian 2004. A unified model of language acquisition. – Judith F. Kroll, Anette de Groot (Eds.). *Handbook of Bilingualism: Psycholinguistic Approaches*. Oxford: Oxford University Press.
- Martin, Maisa 2004. Three structures of Finnish and the Processability Theory. Lena Ekberg, Gisela Håkansson (Eds.). *NORDAND 6. Sjätte konferensen on Nordens språk som andraspråk*. Lund: Lunds universitetet, institutet för nordiska språk, 201–212.
- Martin, Maisa, Sanna Mustonen, Nina Reiman, Marja Seilonen 2010. On becoming an independent user. – Inge Bartning, Maisa Martin, Ineke Vedder (Eds.). *Communicative Proficiency and Linguistic Development: Intersections between SLA and Language Testing Research*. *Eurosla Monographs Series 1*. European Second Language Association, 57–81.
- Masini, Francesca 2005. Multi-word expressions between syntax and the lexicon: The case of Italian verb-particle constructions. – *SKY Journal of Linguistics* 18, 145–173.
- Matsak, Erika 2005. Dialogue system for extracting Logic constructions in natural language texts. – Hamid R. Arabnia, Rose Joshua (Eds.). *Proceedings of the 2005 International Conference on Artificial Intelligence, ICAI 2005, Las Vegas, Nevada, USA, June 27–30, 2005, Volume 2*. CSREA Press, 791–797.
- Matsak, Erika 2009a. *Discovering Logical Constructs from Estonian Children Language*. Thesis on Informatics and System Engineering C51 Tallinn University of Technology. Tallinn: TTÜ Kirjastus.
- Matsak, Erika 2009b. Representing logical inference steps with digital circuits. – *Lecture Notes in Artificial Intelligence, HCI International 2009. Lecture Notes in Computer Science 5618*. Berlin, Heidelberg: Springer, 178–184.
- Metslang, Helle, Ingrid Krall, Renate Pajusalu, Kristi Saarlo, Elle Sõrmus, Silvi Vare 2003. *Keelehärm*. Tallinn: Tallinna Pedagoogikaülikooli Kirjastus.
- Muischnek, Kadri 2006. *Verbi ja noomeni püsiühendid eesti keeles*. *Dissertationes philologiae Estonicae Universitatis Tartuensis 17*. Tartu: Tartu Ülikooli Kirjastus.
- Möhle, Dorothea, Manfred Raupach 1989. Language transfer of procedural knowledge. – Hans W. Dechert, Michael R. Raupach (Eds.). *Transfer in Language Production*. Norwood, New Jersey: Ablex Publishing Corporation, 195–216.
- Myles, Florence 2007. Using electronic corpora in SLA research. – Dalila Ayoun (Ed.). *French Applied Linguistics. Language Learning and Language Teaching 16*. Amsterdam: John Benjamins, 377–400.

- Nation, Paul, Robert Waring 1997. Vocabulary size, text coverage and word lists. – Norbert Schmitt, Michael McCarthy (Eds.). *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press, 6–19. <http://www.robwaring.org/papers/cup.html> (14.11.2009).
- Nesselhauf, Nadja 2005. *Collocations in a Learner Corpus*. Studies in Corpus Linguistics 14. Amsterdam, Philadelphia: John Benjamins.
- Nielsen, Sandro 2002. *Lexicographical Basis for an Electronic Bilingual Accounting Dictionary: Theoretical Considerations*. A translation and revised version of an original paper in Danish, which was published in *LexicoNordica* 9–2002, 173–194. <http://www.sprog.asb.dk/sn/lexicographicalbasis.htm> (10.05.2010).
- Oakes, Michael 1998. *Statistics for Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Orpin, Debbie 2005. *Corpus linguistics and critical discourse analysis: Examining the ideology of sleaze*. – *International Journal of Corpus Linguistics* 10 (1), 37–61. doi:[10.1075/ijcl.10.1.03orp](https://doi.org/10.1075/ijcl.10.1.03orp)
- Pajusalu, Renate, Ilona Tragel 2007. *Word and construction as units of categorization: The case of change predicates in Estonian*. – Andrea C. Schalley, Drew Khlentzos (Eds.). *Mental States, Vol. 2: Language and Cognitive Structure*. Studies in Language Companion Series 93. Amsterdam: John Benjamins, 289–310.
- Paomees, Eve 2003. *Suulise kõne ilmingud 2000. aasta riigiexsamikirjandites*. Bakalaureusetöö. Käsikiri Tartu Ülikooli eesti keele osakonnas.
- Pastuhhova, Olga 2007. *Vene emakeelega üliõpilaste eesti keele verbirektiooni omandamisest*. Magistritöö. Käsikiri Tartu Ülikooli eesti ja soome-ugri keeleteaduse osakonnas. [http://dspace.utlib.ee/dspace/bitstream/10062/2933/1/pastuhhova\\_olga.pdf](http://dspace.utlib.ee/dspace/bitstream/10062/2933/1/pastuhhova_olga.pdf) (21.12.2009).
- Pavlenko, Anna, Scott Jarvis 2002. *Bidirectional transfer*. – *Applied Linguistics* 23 (2), 190–214. doi:[10.1093/applin/23.2.190](https://doi.org/10.1093/applin/23.2.190)
- Petch-Tyson, Stephanie 1998. *Writer/reader visibility in EFL written discourse*. – Sylviane Granger (Ed.). *Learner English on Computer*. London, New York: Longman, 107–118.
- Pieneman, Manfred 1998. *Language Processing and Second Language Development*. Processability Theory. Studies in Bilingualism 15. Amsterdam: John Benjamins.
- Pienemann, Manfred 2005. *Crosslinguistic Aspects of Processability Theory*. Studies in Bilingualism 30. Philadelphia: John Benjamins.

- Pool, Raili 2007. Eesti keele teise keelena omandamise seaduspärasusi täis- ja osasihitise näitel. *Dissertationes philologiae Estonicae Universitatis Tartuenssis* 19. Tartu: Tartu Ülikooli Kirjastus.
- Puhtvend, Kädi 2009. I–IV klassi laste kujundlik keelekasutus. Õpetamine ja oskused. Bakalaureusetöö. Käsikiri Tallinna Ülikooli eesti keele ja kultuuri instituudis.
- Raamdokument 2007 = Euroopa keeleõppe raamdokument: õppimine, õpetamine, hindamine. Tartu: Haridus- ja Teadusministeerium, 2007.
- Rayson, Paul 2002. *Matrix: A Statistical Method and Software Tool for Linguistic Analysis through Corpus Comparison*. Ph.D. thesis in Computer Science. Computing Department Lancaster University. <http://ucrel.lancs.ac.uk/people/paul/publications/phd2003.pdf> (30.04.2009).
- Reppen, Randi, Susan M. Fitzmaurice, Conrad D. S. Biber (Eds.) 2002. *Using Corpora to Explore Linguistic Variation*. *Studies in Corpus Linguistics* 9. Amsterdam, Philadelphia: John Benjamins.
- Ringbom, Håkan 1998. Vocabulary frequencies in advanced learner English: A cross-linguistic approach. – Sylviane Granger (Ed.). *Learner English on Computer*. London, New York: Longman, 41–52.
- Ringbom, Håkan 2007. *Cross-Linguistic Similarity in Foreign Language Learning*. Clevedon: Multilingual Matters LTD.
- Römer, Ute 2005. *Progressives, Patterns, Pedagogy: A Corpus-driven Approach to English Progressive Forms, Functions, Contexts and Didactics*. *Studies in Corpus Linguistics* 18. Amsterdam: John Benjamins.
- Römer, Ute 2007. Learner language and the norms in native corpora and EFL teaching materials: A case study of English conditionals. – Sabine Volk-Birke, Julia Lippert (Eds.). *Anglistentag 2006 Halle. Proceedings*. Trier: Wissenschaftlicher Verlag Trier, 355–363.
- Saareväli, Tatjana-Elis 2005. Verbikasutuse probleeme gümnaasiumilõpetajate kirjalikes töödes. Bakalaureusetöö. Käsikiri Tallinna Ülikooli eesti keele ja kultuuri instituudis.
- Sahkai, Heete 2006. Konstruktsioonipõhise keelekirjelduse võimalustest adessiivse viisi- ja põhjussmääruse näitel. – *Keel ja Kirjandus* 10, 816–831.
- Sahkai, Heete 2008. Konstruktsioonipõhine keelemudel ja sõnaraamatumudel. – *Eesti Rakenduslingvistika Ühingu aastaraamat* 4, 177–186.
- Saukkonen, Pauli, Marjatta Haipus, Antero Niemikorpi, Helena Sulkala 1979. *Suomen kielen taajuusanasto*. Porvoo, Helsinki, Juva: Söderström.
- Scott, Mike, Christopher Tribble 2006. *Textual Patterns: Key Words and Corpus Analysis in Language Education*. Philadelphia: John Benjamins.

- Sepper, Maria-Maren 2006. Indirektaal eesti 19. sajandi lõpu ja 20. sajandi aja- ja ilukirjanduses. Magistritöö. Käsikiri Tallinna Ülikooli eesti keele ja kultuuri instituudis.
- Sinclair, John 2004. *Trust the Text: Language, Corpus and Discourse*. London: Routledge.
- Stubbs, Michael 1996. *Text and Corpus Analysis*. Oxford: Blackwell Publ.
- Stubbs, Michael 2001. *Words and Phrases. Corpus Studies of Lexical Semantics*. Oxford: Blackwell Publ.
- Summers, Della 1996. Computer lexicography: The importance of representativeness in relation to frequency. – Jenny Thomas, Michael H. Short (Eds.). *Using Corpora for Language Research*. London, New York: Longman, 260–266.
- Taylor, Charlotte 2008. What is corpus linguistics? What the data says. – *ICAME Journal* 32, 179–200. [http://icame.uib.no/ij32/ij32\\_179\\_200.pdf](http://icame.uib.no/ij32/ij32_179_200.pdf) (24.03.2009).
- Tognini-Bonelli, Elena 2001. *Corpus Linguistics at Work. Studies in Corpus Linguistics* 6. Amsterdam, Philadelphia: John Benjamins.
- Tognini-Bonelli, Elena 2002. Functionally complete units of meaning across English and Italian: Towards a corpus-driven approach. – Bengt Altenberg, Sylviane Granger (Eds.). *Lexis in Contrast. Corpus-based Approaches. Studies in Corpus Linguistics* 7. Philadelphia: John Benjamins, 73–95.
- Tomasello, Michael 2006. Acquiring linguistic constructions. – Deanna Kuhn, Robert Siegler (Eds.). *Handbook of Child Psychology, Vol. 2, Cognition, Perception, and Language*. 6th edition. New York: Wiley.
- Tono, Yukio, Megumi Aoki 1998. Developing the optimal learning list of irregular verbs based on the native and learner corpora. – International symposium on computer learner corpora, second language acquisition and foreign language teaching. The Chinese University of Hong Kong, 14–16 December, 1998. <http://leo.meikai.ac.jp/~tono/paper/handout.pdf> (5.07.2008).
- Tooding, Liina-Mai 1999. *Andmeanalüüs sotsiaalteadustes*. Tartu: Tartu Ülikooli Kirjastus.
- Torn, Reeli 2003. Object case marking: English influence on the use of Estonian by three Estonian-English bilingual children. – *Kalbotyra/Linguistics* 53 (3), 126–134.
- Torn, Reeli 2004. Sihitisekäänete ja tulevikuvormide kasutamine kolme eesti-inglise kakskeelse lapse eesti keeles. – Birute Klaas, Silvi Tenjes (Toim.). *Emakeel ja teised keeled IV. Tartu Ülikooli eesti keele (võorkeelena) õppetooli toimetised* 3. Tartu: Tartu Ülikooli Kirjastus, 177–191.

- Vanem, Lilian 2006. Eesti ühendverbide vasted vene keeles. Sõnastik ja didakti-  
lisi soovitusi. Magistritöö. Käsikiri Tallinna Ülikooli eesti keele ja kultuuri  
instituudis.
- Vare, Silvi 2001. Acquisition of Estonian as a second language. – Jüri Valge, Raimo  
Raag (Koost.). 100 aastat akadeemilist eesti keele õpet Uppsala Ülikoolis.  
Eesti keele ja keelepoliitika ettekanded. Tartu, Uppsala: Eesti Haridusmi-  
nisteerium, 51–59.
- Vihman, Marilyn May, Maigi Vija 2006. The acquisition of verbal inflection in  
Estonian. – Natalia Gagarina, Insa Gülzow (Eds.). The Acquisition of Verbs  
and their Grammar: The Effect of Particular Languages. Studies in Theore-  
tical Psycholinguistics 33. Dordrecht: Springer, 263–295.
- Vija, Maigi 2007. Pronoomenid lapsekeeles: mõnda *mina* ja *sina* omandamisest. –  
Eesti Rakenduslingvistika Ühingu aastaraamat 3, 373–384.
- Vill, Taimi 2001. Viga eesti keeles eesti kooliõpilase hinnangul. Bakalaureusetöö.  
Käsikiri Tartu Ülikooli eesti keele osakonnas.
- Wilson, Stephen 2003. Lexically specific constructions in the acquisition  
of inflection in English. – Child Language 30, 75–115. doi:[10.1017/  
S0305000902005512](https://doi.org/10.1017/S0305000902005512)
- Õim, Haldur, Heili Orav, Piia Taremaa 2009. Lihtlause semantika: teoreetiline  
kontseptsioon ja arvutianalüüsi võimalused. – Keel ja Kirjandus 7, 489–  
504.
- Õim, Katre 2009. Alternatiivseid mooduseid fraseoloogia esitamiseks sõnasti-  
kus. – Pille Eslon, Katre Õim (Toim.). Korpusuuringute metodoloogia ja  
märgendamise probleemid. Tallinna Ülikooli eesti keele ja kultuuri insti-  
tuudi toimetised 11. Tallinn: Tallinna Ülikooli Kirjastus, 136–164.
- Östman, Jan-Ola, Mirjam Fried (Eds.) 2005. Construction Grammars: Cognitive  
Grounding and Theoretical Extensions. Constructional Approaches to  
Language 3. Amsterdam, Philadelphia: John Benjamins.

**Pille Eslon, Katre Õim, Annekatrin Kaivapalu, Reili Argus, Erika Matsak**

Tallinna Ülikool, eesti keele ja kultuuri instituut

Narva mnt 29

10120 Tallinn, Estonia

[pille.eslon@tlu.ee](mailto:pille.eslon@tlu.ee)

[katre.oim@tlu.ee](mailto:katre.oim@tlu.ee)

[kaivapa@tlu.ee](mailto:kaivapa@tlu.ee)

[reili.argus@tlu.ee](mailto:reili.argus@tlu.ee)

[matsak@tlu.ee](mailto:matsak@tlu.ee)

## **Perspectives of the research of the first and of the second language acquisition**

PILLE ESLON, KATRE ÕIM,  
ANNEKATRIN KAIVAPALU,  
REILI ARGUS, ERIKA MATSAK  
Tallinn University

The main objective of the research is to compare the linguistic development of first (L1) and second language (L2) learners from a dual perspective: 1) stages of acquisition (young children – pre-schoolers – primary school pupils – basic school pupils – upper-secondary school students – adults) and 2) levels of proficiency (A1 – A2 – B1 – B2 – C1 – C2). The intersection of these two coordinate systems forms the general framework on the basis of which one can describe the written and oral language use and linguistic development of L1 and L2 learners of Estonian performing similar communicative tasks. The results of the research should give an overview of the similarities and differences in the stages and proficiency levels of language acquisition in the L1 and L2. They should help researchers understand in what order and in what manner language learners create and use new linguistic structures; whether L1 and L2 are acquired in parallel, and if yes, then to what extent this is so; what is the role of the frequency, transparency, grammaticality and acceptability of linguistic elements and (idiomatic) constructions in language acquisition; and to what extent learners rely on rules or analogy in language acquisition.

The methodological objective of the research is to determine the most typical and frequent communicative goals that can be compared in the case of language learners of different ages in terms of acquisition stages and of proficiency levels. Investigation of corpus data will help the researchers who are implementing the research to identify words and word forms, syntactic and logical constructions that are used to achieve similar communicative goals. This aspect of the research is expected to identify a considerable number of idiomatic expressions whose order of acquisition can then be observed through various acquisition stages and on various proficiency levels. It will also be important to study L1 and L2 text pro-

duction and the psycholinguistic mechanisms of language acquisition by means of experiments specifically adapted or designed for the Estonian language.

The research will be based on the Common European Framework of Reference for Languages (CEFR) and the Development Strategy of the Estonian Language. The present descriptions of reference levels of Estonian language proficiency are mostly based on one's teaching experience. The descriptions of reference levels A, B and C1, which have counterparts in the CEFR, are given in practical, 'can-do' terms; the linguistic aspect of language acquisition remains in the background. No corpus-based longitudinal studies to explore morpho-syntactic constructions and lexical patterns characteristic of various proficiency levels of Estonian have yet been conducted in Estonia.

Questions of interest for researchers include the following: whether lexical patterns and grammatical constructions become gradually more complex and varied in active communication, whether their number increases, stabilises or decreases, whether the learner's linguistic creativity increases with age or whether the language use of adults remains similar to that of teenagers. Researchers have also investigated how and in which order a learner chooses and uses words and word forms, what the grammar of linguistic units is, what stages the learner goes through while producing a text, what the characteristics of this process are, etc.

**Keywords:** corpus linguistics; first and second language acquisition; linguistic description levels of proficiency; morphosyntax; Estonian